**HEART DISEASE PREDICTION USING MACHINE LEARNING**

A Project Synopsis Submitted

in Partial Fulfilment of the Requirements

for the Degree of

**BACHELOR OF TECHNOLOGY**

in

**Computer Science & Engineering**

by

**ISHA SHAHI (Roll No. 1712010042)**

Under the Supervision of

**RAHUL CHAKRAVORTY**

**(Assistant Professor)**



**Submitted to**

**Department of Computer Science & Engineering**

**INSTITUTE OF TECHNOLOGY AND MANAGEMENT , GORAKHPUR**

**Feb , 2021**

# Certificate

Certified that **ISHA SHAHI** has carried out the project work presented in this project entitled

**“HEART DISEASE PREDICTION USING MACHINE LEARNING”**  for the award of “**Bachelor of**

**Technology”** from Dr.**A.P.J Abdul kalam Technical University (AKTU )** , Lucknow

under my supervision. The content of the project do not form the basis for the award of any

other degree to the candidate or to anybody else .

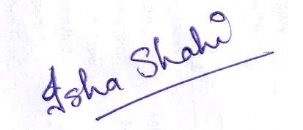
RAHUL CHAKRAVORTY (Project Guide)

Designation : ASSISTANT PROFESSOR

Date : 12 July 2021

# CANDIDATE’S DECLARATION

I declare that this written submission represents my work and ideas in my own words and where other’s ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

****

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Signature)

Isha Shahi

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Name of the student)

1712010042

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Roll No.)

Department of Computer Science & Engineering

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Name of Department)

12/07/2021

Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# APPROVAL SHEET

This project entitled “**HEART DISEASE PREDICTION USING MACHINE LEARNING**” by **ISHA SHAHI** **,** is approved for the degree of Bachelor of Technology.

### **Examiner**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Supervisor (s)**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Head of Department**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Dean, Research & Development or**

**Other Dean/Professor to be nominated**

### **by the Vice Chancellor in his absence**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Date: 15/07/2021

Place: Gorakhpur

# Acknowledgement

We would like to take the opportunity to thank those people who guided us and supported me during this period of my study. Without their valuable contributions, this dissertation work would not have been possible. We, as a team, have worked together on building this project from scratch. A lot of thinking has been done in designing the project before the actual implementation has taken place. We have taken the aid and assistance of different resources available across the internet for gaining the domain knowledge regarding the development of the project. We would like to thank the training and development courses in Udemy from where we learnt Machine Learning. We are grateful to our project guide **RAHUL CHAKRAVORTY**, Assistant Professor, for providing proper guidance and guiding us in the correct direction and always being there as a support system. Also, we would like to thank our friends who have supported us and provided assistance during the development of this project. Lastly, we would remiss in our duties if we do not thank our parents who have been a constant moral support.

We are thankfully to Dr. Shachi Mall (HOD,CSE) and Miss Pritee Yadav for their support.

We express our gratitude and thanks to all faculties and staff members of Computer Science

and Engineering department for their sincere cooperation in furnishing relevant information

to compute this project well in time successfully.

**Group 7 (ISHA SHAHI)**

**Roll No- (1712010042 )**

**Date: 12 JULY 2021**

**Place: Gorakhpur**

# Abstract

In 21st century, Prosperity is all about being successful in career, earning money and leading a comfortable life. The most important element which is overlooked in this competitive world is Health. People have been so prone to various diseases. This has been due to the courtesy of pollution, ill eating habits and lack of exercise. One of the most deadliest disease affecting the human race is those related to the Heart. There have been a massive rise in case of Cardiac Arrest and Heart Attacks. It’s all due to modern lifestyle and food habits that we boast of. This paper proposes a set of idea which predicts the possibility of heart disease and provides preventive measures in form of diet plan. The prediction is done on the basis of user’s medical data through Data Mining. The user have to input the reports of regular checkup. Various data from the reports are taken and the prediction is fully data driven and it doesn’t require any infrastructure which makes it hassle-free and that is the benefit over existing systems. The idea is provided to user through Website. Many a times it becomes increasingly difficult to detect heart problems early on, especially those related to blockage of the arteries. Modern lifestyle has led to sudden heart problems, leading to fatal heart attacks for the younger generation. The prediction of any heart related anomalies, can not only prevent casualties related to heart, but also help to correctly predict the exact disease and help the patient get the appropriate treatment.

Table of Contents

**S.NO. TITLE PAGE**

1 [Certificate 2](#_Toc78621375)

2 [CANDIDATE’S DECLARATION 3](#_Toc78621376)

3 [APPROVAL SHEET 4](#_Toc78621377)

4 [Acknowledgement 5](#_Toc78621380)

5 [Abstract 6](#_Toc78621381)

6 [Table of Contents 7](#_Toc78621382)

7 [LIST OF FIGURES 9](#_Toc78621383)

**CHAPTER- 1** [**INTRODUCTION 10**](#_Toc78621385)

[1.1 Background 11](#_Toc78621386)

[1.2 Machine Learning 11](#_Toc78621387)

[1.2.1 Supervised Learning 11](#_Toc78621388)

[1.2.2 Unsupervised Learning 11](#_Toc78621389)

[1.2.3 Reinforcement Learning 11](#_Toc78621390)

**CHAPTER -2** [**CASE STUDY 13**](#_Toc78621392)

**CHAPTER -3** [**LITERATURE SURVEY 15**](#_Toc78621394)

**CHAPTER -4** [**IMPLEMENTATION 21**](#_Toc78621396)

[4.1 Requirement Analysis 21](#_Toc78621397)

[4.1.1 Hardware Requirement 21](#_Toc78621398)

[4.1.2 Software Requirement 21](#_Toc78621399)

[4.2 Non-Functional Requirement 21](#_Toc78621400)

[4.2.1 Security 21](#_Toc78621401)

[4.2.2 Usability 21](#_Toc78621402)

[4.2.3 Portability 21](#_Toc78621403)

[4.3 About Data Sets 21](#_Toc78621404)

[4.3.1 Data 22](#_Toc78621405)

[4.3.2 Importing Libraries 22](#_Toc78621406)

[4.3.3 Handling missing values 22](#_Toc78621407)

[4.4 Algorithm 23](#_Toc78621408)

**CHAPTER- 5** [**METHODOLOGY 26**](#_Toc78621410)

**CHAPTER- 6** [**SCREENSHOTS 28**](#_Toc78621412)

**CHAPTER- 7** [**CONCLUSION AND FUTURE SCOPE 39**](#_Toc78621415)

[7.1 Conclusion 39](#_Toc78621416)

[7.2 Future Scope 39](#_Toc78621417)

**CHAPTER- 8** [**REFRENCES 40**](#_Toc78621419)

# LIST OF FIGURES

[Figure 1-FLOWCHART 27](file:///C:\Users\dell\OneDrive\Desktop\Report%20hdp.docx#_Toc77503927)

[Figure 2-SCREENSHOT 1 28](#_Toc77503928)

[Figure 3-SCREENSHOT-2 29](#_Toc77503929)

[Figure 4-SCREENSHOT-3 30](file:///C:\Users\dell\OneDrive\Desktop\Report%20hdp.docx#_Toc77503930)

[*Figure 5-SCREENSHOT-4* 31](#_Toc77503931)

[Figure 6-SCREENSHOT-5 32](#_Toc77503932)

[Figure 7-SCREENSHOT-6 33](#_Toc77503933)

[Figure 8-SCREENSHOT -7 34](#_Toc77503934)

[Figure 9-SCREENSHOT-8 34](#_Toc77503935)

[Figure 10-SCREENSHOT-9 35](#_Toc77503936)

[Figure 11-SCREENSHOT-10 35](file:///C:\Users\dell\OneDrive\Desktop\Report%20hdp.docx#_Toc77503937)

[Figure 12-SCREENSHOT-11 36](#_Toc77503938)

[Figure 13-SCREENSHOT-12 36](#_Toc77503939)

[Figure 14-SCREENSHOT-13 37](file:///C:\Users\dell\OneDrive\Desktop\Report%20hdp.docx#_Toc77503940)

[Figure 15-SCREENSHOT-14 37](#_Toc77503941)

[Figure 16-SCREENSHOT-15 38](#_Toc77503942)

# Chapter 1

# Introduction

Machine Learning is the field of study where computers have the ability to learn without

being explicitly programmed. It is one of the most exciting technologies that one would

have ever come across. As it is evident from the name, it gives the computer that makes

it more similar to humans: The ability to learn. Machine learning is actively being used

today, perhaps in many more places than one would expect. The healthcare industry is

one of them. Machine Learning plays an essential role in predicting presence/absence of

Loco-motor disorders, Heart diseases and more. Such information, if predicted well in

advance, can provide important insights to doctors who can then adapt their diagnosis

and treatment per patient basis.

In today’s world cardiac arrest is one of the most significant causes of mortality. Predicting heart

diseases is a challenge for the healthcare department as it is hard to narrow

down the specific factors which contribute to deteriorating heart condition. Some of

the factors as seen from the past data are diabetes, age, high cholesterol, abnormally

high pulse rate, high blood pressure and etc. We collected data relevant to this field

containing all attributes related to it. Meaningful knowledge was retrieved from large

dataset by data mining. By using the concepts of statistics we understood and analyzed

the patterns present in the data.

In this study we classified the patients in only two categories namely:

1) Has heart disease or

2) Does not have heart disease by representing them in integers 1 and 0 respectively.

Logistic Regression ,Random Forest and Naive Bayes are the algorithms we used which

predicted with the highest accuracies .Logistic Regression classifies observations using

a complex cost function called the Sigmoid function. When large number of individual

trees work as an ensemble then it is known as Random Forest classifier .All the trees predict

different classes for the patients and the prediction with the most votes is returned

by Random Forest. Naive Bayes is a probabilistic model that is based on the Bayes

Theorem.

# 1.1Background

Heart disease affects millions of people, and it remains the chief cause of death in the world. Medical diagnosis should be proficient, reliable, and aided with computer techniques to reduce the effective cost for diagnostic tests. Data mining is a software technology that helps computers to build and classify various attributes. This research paper uses classification techniques to predict heart disease. This section gives a portrayal of the related subjects like machine learning and its methods with brief descriptions, data pre-processing, evaluation measurements and description of the dataset used in this research.

## 1.2Machine Learning

Machine learning is an emerging subdivision of artificial intelligence. Its primary focus is to design systems, allow them to learn and make predictions based on the experience. It trains machine learning algorithms using a training dataset to create a model. The model uses the new input data to predict heart disease. Using machine learning, it detects hidden patterns in the input dataset to build models. It makes accurate predictions for new datasets. The dataset is cleaned and missing values are filled. The model uses the new input data to predict heart disease and then tested for accuracy. Machine learning techniques are classified as:

## 1.2.1Supervised Learning

The model is trained on a dataset that is labelled. It has input data and its outcomes. Data are classified and split into training and test dataset. Training dataset trains our model while testing dataset functions as new data to get accuracy of the model. The dataset exists with models and its output. The classification and regression are its example.

## 1.2.2Unsupervised Learning

Data used to train are not classified or labelled in the dataset. Aim is to find hidden patterns in the data. The model is trained to develop patterns. It can easily predict hidden patterns for any new input dataset, but upon exploring data, it draws conclusion from datasets to describe hidden patterns. In this technique, no responses in the dataset are seen. The clustering method is an example of an unsupervised learning technique.

## 1.2.3Reinforcement Learning

It does not use labelled dataset nor the results are associated with data, thus model learns from the experience. In this technique, the model improves its presentation based on its association with environment and figures out how to discuss its faults and to get the right outcome through assessment and testing various prospects.

# Chapter 2

# Case Study

Heart disease is the biggest cause of death of both men and women is USA, England and

China. It is the reason of death in every 4 out of 10 deaths. A person dies of heart attack

every minute in USA.

According to stats, heart disease mostly occurs in adults over the age of 35. Many

reports suggests that people living in middle and low income countries are more prone

to heart disease. There are several types of heart disease but the three most commonly

known are Coronary Artery Disease, Congestive Heart Failure and Bad Heart Rhythms.

Coronary Artery Disease: This is a problem with blood vessels that delivers blood

to the heart muscles. If the blood vessels gets very small, or if they get blocked, blood

cannot flow through them normally. As less blood is supplied to the heart muscles, the

muscles cannot work at their normal capacity. The heart muscles can become weak.

Heart muscle may die if the blood flow stops. Blocked arteries in the heart are often

caused by smoking, high cholesterol, high blood pressure and inherited traits from parents. All

these problems damage the lining of heart’s blood vessels and make them

narrowed or blocked completely. Heart disease present in different age groups.

Congestive Heart Failure: This is a condition which means that the heart is not

pumping at its normal level. Two common causes for this are a weak heart muscle and

abnormal heart valves. The valves might not allow enough blood through because they

are too narrow. Or the valve might ”leak” and let blood flow back inside the heart.

When the heart valves do not work normally, the heart muscle has to do extra work and

it may get tired.

Bad Heart Rhythms: This is a problem with the electrical activity in the heart. This

can make the heart beat too fast or too slow. Bad Heart Rhythms might make the

heart stop pumping blood. The heart needs a normal rhythm to pump the blood well.

If the rhythm is too fast, the heart might not have enough time for blood to enter the

chambers, so there would not be enough blood moving through the heart with each beat.

If the heart is too slow, there might not be enough contractions of the heart to supply

the body with the blood that it needs.

The most commonly observed symptoms in people suffering from heart disease are:

* Pain in the chest–the heart muscle does not get enough flow to keep it going.
* Trouble Breathing–blood might back up into the lungs.
* Palpitations (a feeling that the heart is beating too fast, too hard, or not regularly).
* Swelling of feet or the legs–blood is backing up from the heart into the lower body.
* Feeling weak because the body and brain are not getting enough blood to supply them

with oxygen.

* Cyanosis (skin turning into blue color) which means that too little oxygen is there in

the bloodstream to supply to the cells in the body.

Congenital Heart Disease refers to a form of heart problem that starts before the

birth (congenital). Congenital heart diseases include a large number of conditions. The

formation of heart before birth is very complex and might be abnormal. Like if part of

the main artery that comes from the heart is narrowed, then it is called coarctation of

the aorta. There might be holes inside the heart which keeps the blood from flowing

normally inside the heart. Other congenital heart problems are caused by abnormal

heart valves. In many cases congenital heart problems are discovered at birth. Other

times the problems might not be detected until and unless the person is a grown adult.

# Chapter 3

# Literature Survey

1. Bo Jin, Chao Che et al. (2018) proposed a “Predicting the Risk of Heart Failure With EHR Sequential Data Modelling” model designed by applying neural network. This paper used the electronic health record (EHR) data from real-world datasets related to congestive heart disease to perform the experiment and predict the heart disease before itself. We tend to used one-hot encryption and word vectors to model the diagnosing events and foretold coronary failure events victimization the essential principles of an extended memory network model. By analysing the results, we tend to reveal the importance of respecting the sequential nature of clinical records [1].
2. Aakash Chauhan et al. (2018) presented “Heart Dy678isease Prediction using Evolutionary Rule Learning”. This study eliminates the manual task that additionally helps in extracting the information (data) directly from the electronic records. To generate strong association rules, we have applied frequent pattern growth association mining on patient’s dataset. This will facilitate (help) in decreasing the amount of services and shown that overwhelming majority of the rules helps within the best prediction of coronary sickness [2].
3. Ashir Javeed, Shijie Zhou et al. (2017) designed “An Intelligent Learning System based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection”. This paper uses random search algorithm (RSA) for factor selection and random forest model for diagnosing the cardiovascular disease. This model is principally optimized for using grid search algorithmic program. Two forms of experiments are used for cardiovascular disease prediction. In the first form, only random forest model is developed and within the second experiment the proposed Random Search Algorithm based random forest model is developed. This methodology is efficient and less complex than conventional random forest model. Comparing to conventional random forest it produces 3.3% higher accuracy. The proposed learning system can help the physicians to improve the quality of heart failure detection [3].
4. “Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques” proposed by Senthilkumar Mohan, Chandrasegar Thirumalai et al. (2019) was efficient technique using hybrid machine learning methodology. The hybrid approach is combination of random forest and linear method. The dataset and subsets of attributes were collected for prediction. The subset of some attributes were chosen from the pre-processed knowledge(data) set of cardiovascular disease .After prep-processing , the hybrid techniques were applied and disgnosis the cardiovascular disease [4].
5. K.Prasanna Lakshmi, Dr. C.R.K.Reddy (2015) designed “Fast Rule-Based Heart Disease Prediction using Associative Classification Mining”. In the proposed Stream Associative Classification Heart Disease Prediction (SACHDP), we used associative classification mining over landmark window of data streams. This paper contains two phases: one is generating rules from associative classification mining and next one is pruning the rules using chi-square testing and arranging the rules in an order to form a classifier. Using these phase to predict the heart disease easily [5].
6. M.Satish, et al. (2015) used different Data Mining techniques like Rule based, Decision Tree, Navie Bayes, and Artifical Neural Network. An efficient approach called pruning classification association rule (PCAR) was used to generate association rules from cardiovascular disease warehouse for prediction of Heart Disease. Heart attack data warehouse was used for pre-processing for mining. All the above discussed data mining technique were described [6].
7. Lokanath Sarangi, Mihir Narayan Mohanty, Srikanta Pattnaik (2015) “An Intelligent Decision Support System for Cardiac Disease Detection”, designed a cost efficient model by using genetic algorithm optimizer technique. The weights were optimized and fed as an input to the given network. The accuracy achieved was 90% by using the hybrid technique of GA and neural networks [7].
8. “Prediction and Diagnosis of Heart Disease by Data Mining Techniques” designed by Boshra Bahrami, Mirsaeid Hosseini Shirvani. This paper uses various classification methodology for diagnosing cardiovascular disease. Classifiers like KNN, SVO classifier and Decision Tree are used to divide the datasets. Once the classification and performance evaluation the Decision tree is examined as the best one for cardiovascular disease prediction from the dataset[8].
9. Mamatha Alex P and Shaicy P Shaji (2019) designed “Prediction and Diagnosis of Heart Disease Patients using Data Mining Technique”. This paper uses techniques of Artificial Neural Network, KNN, Random Forest and Support Vector Machine. Comparing with the above mentioned classification techniques in data mining to predict the higher accuracy for diagnosing the heart disease is Artificial Neural Network[9]
10. Senthil Kumar Mohan et al, proposed Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques in which strategy that objective is to finding critical includes by applying Machine Learning bringing about improving the exactness in the expectation of cardiovascular malady. The expectation model is created with various blends of highlights and a few known arrangement strategies. We produce an improved exhibition level with a precision level of 88.7% through the prediction model for heart disease with hybrid random forest with a linear model (HRFLM) they likewise educated about Diverse data mining approaches and expectation techniques, Such as, KNN, LR, SVM, NN, and Vote have been fairly famous of late to distinguish and predict heart disease[10].
11. Sonam Nikhar et al has built up the paper titled as Prediction of Heart Disease Using Machine Learning Algorithms by This exploration plans to give a point by point portrayal of Naïve Bayes and decision tree classifier that are applied in our examination especially in the prediction of Heart Disease. Some analysis has been led to think about the execution of prescient data mining strategy on the equivalent dataset, and the result uncovers that Decision Tree beats over Bayesian classification system[11].
12. Aditi Gavhane, Gouthami Kokkula, Isha Pandya, Prof. Kailas Devadkar (PhD), Prediction of Heart Disease Using Machine Learning”, In this paper proposed system they used the neural network algorithm multi-layer perceptron (MLP) to train and test the dataset. In this algorithm there will be multiple layers like one for input, second for output and one or more layers are hidden layers between these two input and output layers. Each node in input layer is connected to output nodes through these hidden layers. This connection is assigned with some weights. There is another identity input called bias which is with weight b, which added to node to balance the perceptron. The connection between the nodes can be feedforwarded or feedback based on the requirement[12].
13. Abhay Kishore et al, developed Heart Attack Prediction Using Deep Learning in which This paper proposes a heart attack prediction system using Deep learning procedures, explicitly Recurrent Neural System to predict the probable prospects of heart related infections of the patient. Recurrent Neural Network is a very ground-breaking characterization calculation that utilizes Deep Learning approach in Artificial Neural Network. The paper talks about in detail the significant modules of the framework alongside the related hypothesis. The proposed model deep learning and data mining to give the precise outcomes least blunders. This paper gives a bearing and point of reference for the advancement of another type of heart attack prediction platform. Prediction stage[13].
14. Lakshmana Rao et al, Machine Learning Techniques for Heart Disease Prediction in which the contributing elements for heart disease are more (circulatory strain, diabetes, current smoker, high cholesterol, etc..). So, it is difficult to distinguish heart disease. Different systems in data mining and neural systems have been utilized to discover the seriousness of heart disease among people. The idea of CHD ailment is bewildering, in addition, in this manner, the disease must be dealt with warily. Not doing early identification, may impact the heart or cause sudden passing. The perspective of therapeutic science furthermore, data burrowing is used for finding various sorts of metabolic machine learning a procedure that causes the framework to gain from past information tests, models without being expressly customized. Machine learning makes rationale dependent on chronicled information[14].
15. Mr. Santhana Krishnan.J and Dr. Geetha.S, Prediction of heart disease using machine learning algorithm This Paper predicts heart disease for Male Patient using Classification Techniques. The detailed information about Coronary Heart diseases such as its Facts, Common Types, and Risk Factors has been explained in this paper. The Data Mining tool used is WEKA (Waikato Environment for Knowledge Analysis), a good Data Mining Tool for Bioinformatics Fields. The all three available Interface in WEKA is used here; Naive Bayes, Artificial Neural Networks and Decision Tree are Main Data Mining Techniques and through this techniques heart disease is predicted in this System. The main Methodology used for prediction is Decision Trees like CART, C4.5, CHAID, J48, ID3 Algorithms, and Naive Bayes Techniques[15].
16. Avinash Golande et al, proposed Heart Disease Prediction Using Effective Machine Learning Techniques in which Specialists utilize a few data mining strategies that are available to support the authorities or doctors distinguish the heart disease. Usually utilized methodology utilized are decision tree, k- closest and Naïve Bayes. Other unique characterization-based strategies utilized are packing calculation, Part thickness, consecutive negligible streamlining and neural systems, straight Kernel selfarranging guide and SVM (Bolster Vector Machine). The following area obviously gives subtleties of systems that were utilized in the examination[16].
17. V.V. Ramalingam et Al, proposed Heart disease prediction using machine learning techniques in which Machine Learning algorithms and techniques have been applied to various medical datasets to automate the analysis of large and complex data. Many researchers, in recent times, have been using several machine learning techniques to help the health care industry and the professionals in the diagnosis of heart related diseases. This paper presents a survey of various models based on such algorithms and techniques and analyse their performance. Models based on supervised learning algorithms such as Support Vector Machines (SVM), KNearest Neighbour (KNN), Naïve Bayes, Decision Trees (DT), Random Forest (RF) and ensemble models are found very popular among the researchers and systems have been applied to different clinical datasets to robotize the investigation of huge and complex information. Numerous scientists, as of late, have been utilizing a few Machine Learning algorithms and techniques have been applied to various medical datasets to automate the analysis of large and complex data. Many researchers, in recent times, have been using several machine learning techniques to help the health care industry and the professionals in the diagnosis of heart related diseases. This paper presents a survey of various models based on such algorithms and techniques and analyze their performance. Models based on supervised learning algorithms such as Support Vector Machines (SVM), KNearest Neighbour (KNN), Naïve Bayes, Decision Trees (DT), Random Forest (RF) and ensemble models are found very popular among the researchers. strategies to enable the wellbeing to mind industry and the experts in the analysis of heart related sicknesses. This paper presents a review of different models dependent on such calculations and methods and analyze their exhibition. Models in light of directed learning calculations, for example, Support Vector Machines (SVM), K- Nearest Neighbour (KNN), Naïve Bayes, Decision Trees (DT), Random Forest (RF) and group models are discovered extremely well known among the scientists[17].
18. Pranav etal. implemented healthcare monitoring system consisting of ECG Sensors AD8232. ECG sensors monitor vital parameters through remote monitoring of patient. The authors have created android application for continuous monitoring of patient ECG. The heart disease is correctly predicted by various data extraction techniques. Data extraction techniques calculate amplitude and RR interval of ECG wave generated by sensor. This technique uses data mining which reduces reduce time and effort of heart disease detection techniques. Authors have also used IoT technology and data mining algorithms for heart disease prediction[18].
19. Shweta Gajbhiye etal. implemented heart attack early prediction using android application. In this system pulse sensors senses the heart rate and oximeter checks level of oxygen in patient body. The data of heart rate and oxygen level received microcontroller will be displayed on LCD module and also send to Wi-Fi module ESP 8266 microcontroller. This microcontroller will further send this data over internet through Wi-Fi. The android application will be used to monitor the heart rate and oxygen level over Wi-Fi[19].
20. Ponugumatla Kalyan etal. , M.Ganesan etal. [13], Avinash Golande etal. [16], Yosuf Amr ElSaadany [20] implemented IoT based heart disease prediction and monitoring system using arduino and raspberry pi 3. In this system AD8232 sensor will measure the heart rate and sent it to arduino board. The arduino board send this data to raspberry pi 3 boards. The GPS receiver Ublox NEO-6Mv2 will detect the position of patient and send it to PL2303 USB to TTL module. This module will USB connectors which alters the received USB information to standard serial port information. This serial port information will be further provided to raspberry pi 3 microcontroller. The data of heart rate beats will be stored over a cloud using HTML and Wi-Fi. Because of this the doctor or other concern user can easily retrieve this information through cloud[20].

# Chapter 4

# Implementation

## 4.1Requirement Analysis

**4.1 Functional Requirement**

### 4.1.1 Hardware Requirement

For PYTHON implementation the sufficient requirements are RAm of 8 GB and Intel

Core i3- 2120CPU@2.0GHz.

### 4.1.2 Software Requirement

We need to install Anaconda, Python and Jupyter Notebook to run the python code on

the Jupyter notebook. We have to install dependencies in Python like numpy, pandas,

matplotlib, scikit-learn, Stats-model, data line.

## 4.2 Non-Functional Requirement

### 4.2.1 Security

Dataset should be loaded securely. Dataset should not be broken or uncleaned.

### 4.2.2 Usability

The scope of the framework is huge. All type of businesses can easily use the system to

enhance their Data Quality.

### 4.2.3 Portability

Since the project we work on is merely an executable file, it is portable. **.**

## 4.3About Data Sets

There is one file heart.csv which has 304 rows. It has missing values which need to be

first filled up with dummy values so that model can be trained and perform better.

### 4.3.1 Data

1. age = age in years

2. sex= (1 = male; 0 = female)

3. cp = chest pain type

4. trestbpsr = testing blood pressure (in mm Hg on admission to the hospital)

5. chol = serum cholestrol in mg/dl

6. fbs = (fasting blood sugar ¿ 120 mg/dl) (1 = true; 0 = false)

7. restecg = resting electrocardiographic results

8. thalach = maximum heart rate achieved

9. exang = exercise induced angina (1 = yes; 0 = no)

10. oldpeak = ST depression induced by exercise relative to rest

11. slope = the slope of the peak exercise ST segment

12. ca = number of major vessels (0-3) colored by flourosopy

13. thal = 3 = normal; 6 = fixed defect; 7 = reversable defect

14. target = 1 or 0

### 4.3.2 Importing Libraries

#importing libraries

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns from sklearn.model selection import train test split

import warnings

warnings.filterwarnings(’ignore’)

### 4.3.3 Handling missing values

The ways to handle the missing values in the dataset are Mean,Median and Most frequently

occuring values.By default the missing values is replaced by ’NaN’ until we

handle it.Here we used the mean method to handle the missing values.

#Taking care of missing values

from sklearn.preprocessing import Imputer

imputer = Imputer(missing values=’NaN’ strategy=’mean’ axis=0)

imputer = imputer.fit(X[:, 1:3])

X[:, 1:3] = imputer.transform(X[:, 1:3])

## 4.4 Algorithm

**Algorithm 1 Logistic Regression**

1: from sklearn.linearmodel import LogisticRegression

2: xdata = df.drop([’target’], axis = 1)

3: y = df.target.values

4: lr = LogisticRegression()

5: lr.fit(xtrain, ytrain)

6: print(’Test Accuracy :.2f .format(lr.score(xtest, ytest)\*100))

#Logistic Regression Test Accuracy 85.25

Test Accuracy 85.25

**Algorithm 2 Naive Bayes**

1: from sklearn.naivebayes import GaussianNB

2: nb = GaussianNB()

3: nb.fit(xtrain, ytrain)

4: print(”NB accuracy: :.2f”.format(nb.score(xtest, ytest)\*100))

# Naive Bayes Test Accuracy

NB accuracy: 85.25

**Algorithm 3 Random Forest**

1: from sklearn.ensemble import RandomForestClassifier

2: rf = RandomForestClassifier(nestimators = 1000, randomstate= 1)

3: rf.fit(xtrain, ytrain)

4: print(”Random Forest accuracy: :.2f”.format(rf.score(xtest, ytest)\*100))

# Random forest Test Accuracy

Random Forest accuracy: 85.25

**Algorithm 4 Support Vector Machine**

1. from sklearn import svm
2. sv = svm.SVC(kernel='linear')
3. sv.fit(X\_train, Y\_train)
4. Y\_pred\_svm = sv.predict(X\_test)

**Algorithm 5 K Nearest Neighbors**

1. from sklearn.neighbors import KNeighborsClassifier
2. knn = KNeighborsClassifier(n\_neighbors=7)
3. knn.fit(X\_train,Y\_train)
4. Y\_pred\_knn=knn.predict(X\_test)

**Algorithm 6 Decision Tree**

1. from sklearn.tree import DecisionTreeClassifier
2. max\_accuracy = 0
3. for x in range(200):
4. dt = DecisionTreeClassifier(random\_state=x)
5. dt.fit(X\_train,Y\_train)
6. Y\_pred\_dt = dt.predict(X\_test)
7. current\_accuracy = round(accuracy\_score(Y\_pred\_dt,Y\_test)\*100,2)
8. if(current\_accuracy>max\_accuracy):
9. max\_accuracy = current\_accuracy
10. best\_x = x
11. #print(max\_accuracy)
12. #print(best\_x)
13. dt = DecisionTreeClassifier(random\_state=best\_x)
14. dt.fit(X\_train,Y\_train)
15. Y\_pred\_dt = dt.predict(X\_test)

# Chapter 5

# Methodology

In order to perform an analysis we collected a dummy dataset from the internet .The

dataset table contains 14 columns , 13 input columns and one output column. Out of

these 14 columns one column takes float value whereas the rest take integer values.Firstly,

we handle the missing values with NaN using the imputer function. Secondly, on performing the

age wise analysis we found out that people of age above 50 are more prone

to having heart diseases. Out of the total patients having heart disease , 159 patients

are above the age of 55, 128 are mid aged patients of the age ranging from 40-54 and

the rest are young ones below 40. In a percentage wise analysis a total of 54.46 per cent

have heart disease and out of this 31.65 per cent patients are female and 65.32 per cent

patients are male. Now at the end during the model building we divide our dataset into

training and testing dataset.The dataset is split in the ratio of 75:25(where 75 per cent

data is used for training the machine and the rest 25 per cent is used for testing purpose) by

using train test split method in sklearn. cross validation module. Then we use

a total of five algorithms,first one being logistic regression(imported from sklearn.linear

model) which gives us a accuracy of 85.25per cent Second, KNN model(imported from

sklearn.neighbors) which gives an accuracy of 43.93 per cent. Third, support vector

method(imported from sklearn.svm) which gives an accuracy of 59.02 per cent Fourth,

we use Naive Bayes algorithm(imported from sklearn.naive bayes) which gives an accuracy of

85.25 per cent. Lastly, we use the random forest classifier(imported from

sklearn.ensemble) which provides us with an accuracy of 85.25 per cent So, we see that

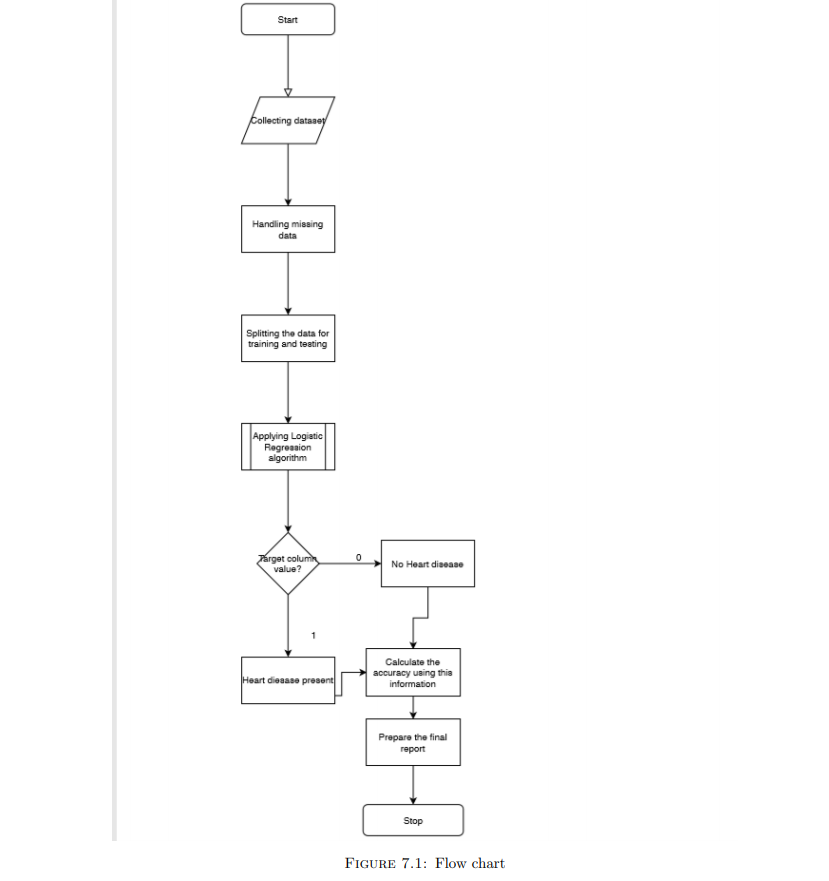
Logistic Regression , Naive Bayes and Random Forest provides us with the best accuracy

out of the five algorithms. So we have decided to use Logistic Regression algorithm. The

Target column contains either 0 or 1. 1 indicates that the patient has heart disease and

0 indicates that the patient does not have a heart disease.

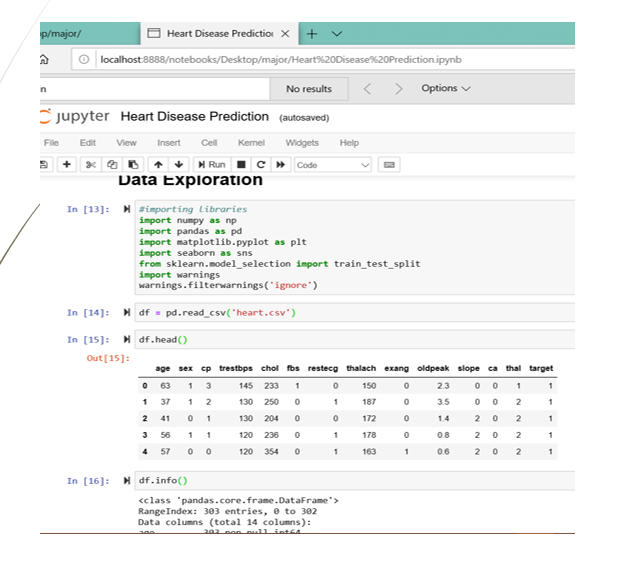
**Figure 1-FLOWCHART**



## 

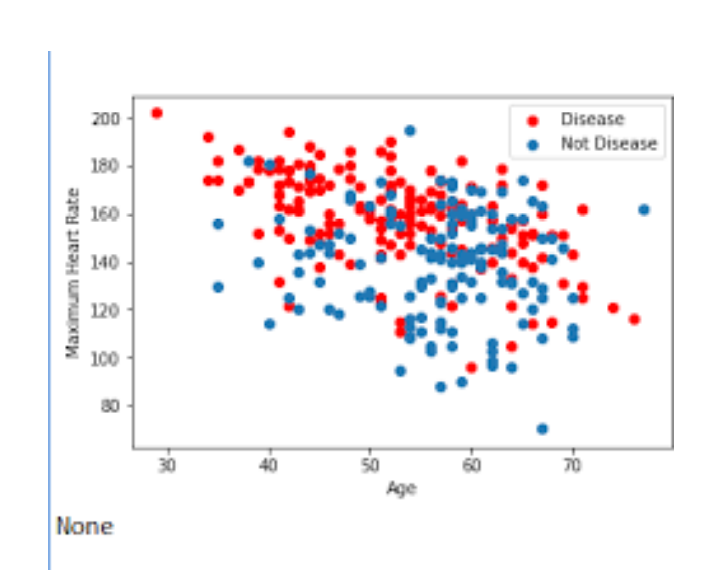
# Chapter 6

# Screenshots

****

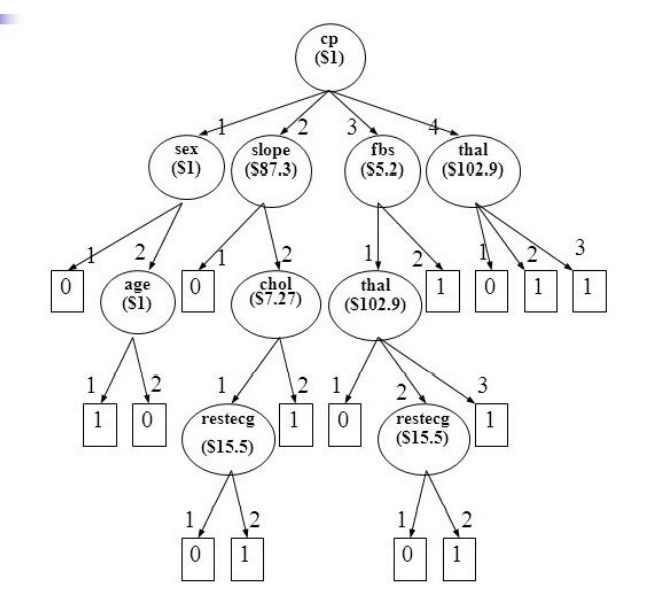
**Figure 2-SCREENSHOT 1**

## 

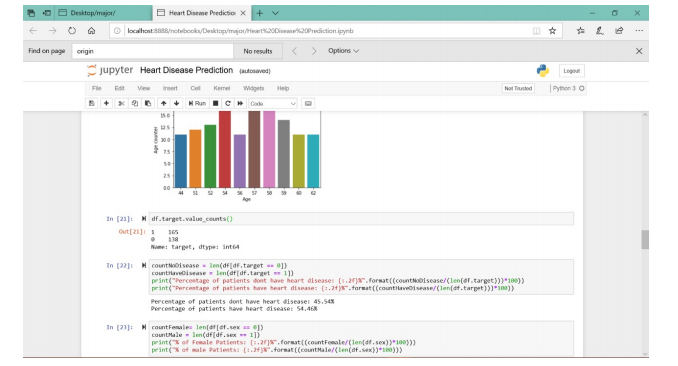


**Figure 3-SCREENSHOT-2**

## 

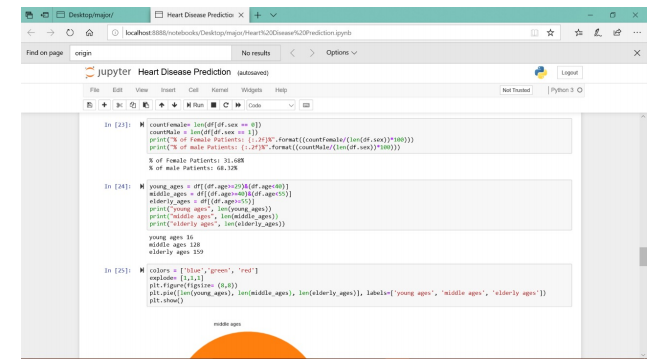


**Figure 4-SCREENSHOT-3**



## 

## ***Figure 5-SCREENSHOT-4***

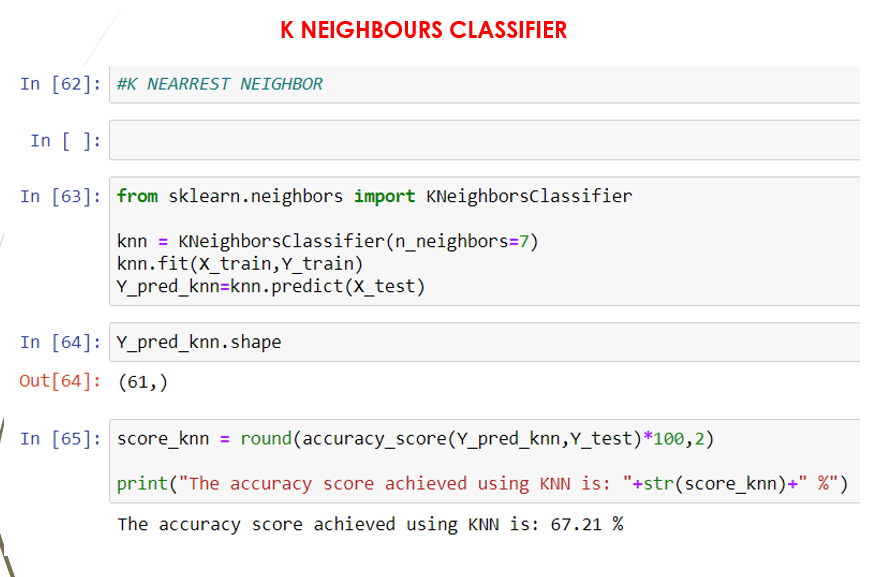


**Figure 6-SCREENSHOT-5**

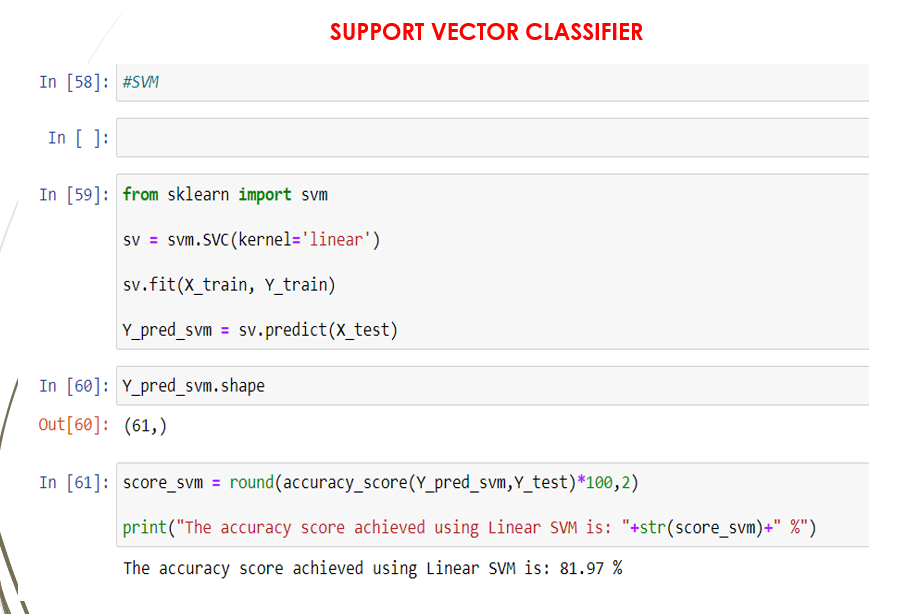
## 



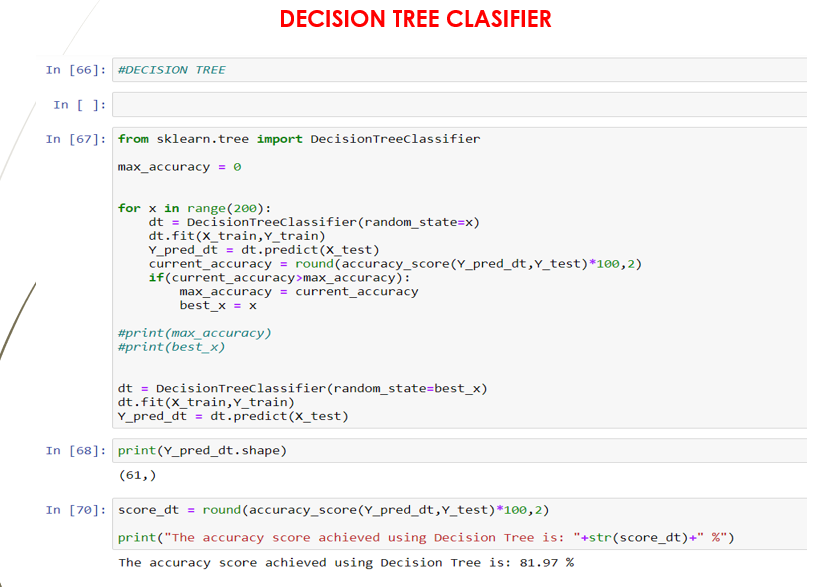
**Figure 7-SCREENSHOT-6**



**Figure 8-SCREENSHOT -7**



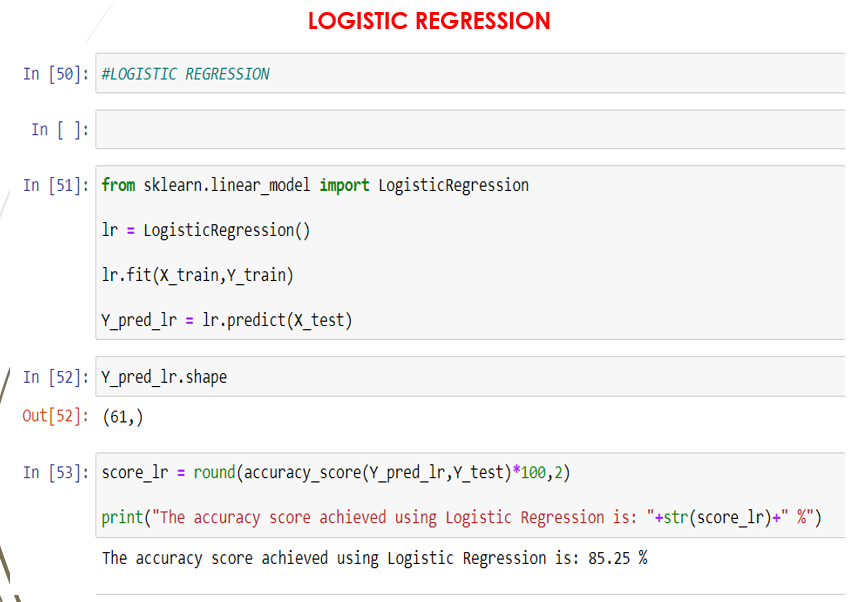
**Figure 9-SCREENSHOT-8**



**Figure 10-SCREENSHOT-9**



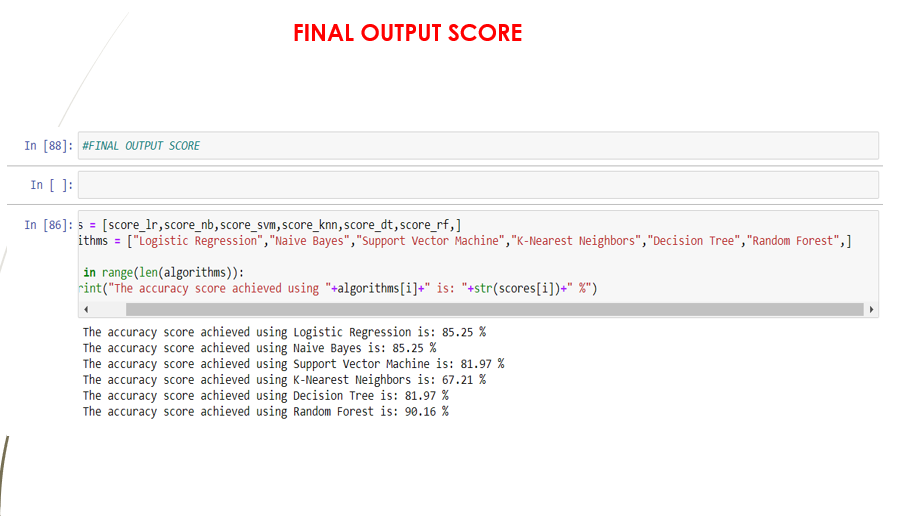
**Figure 11-SCREENSHOT-10**



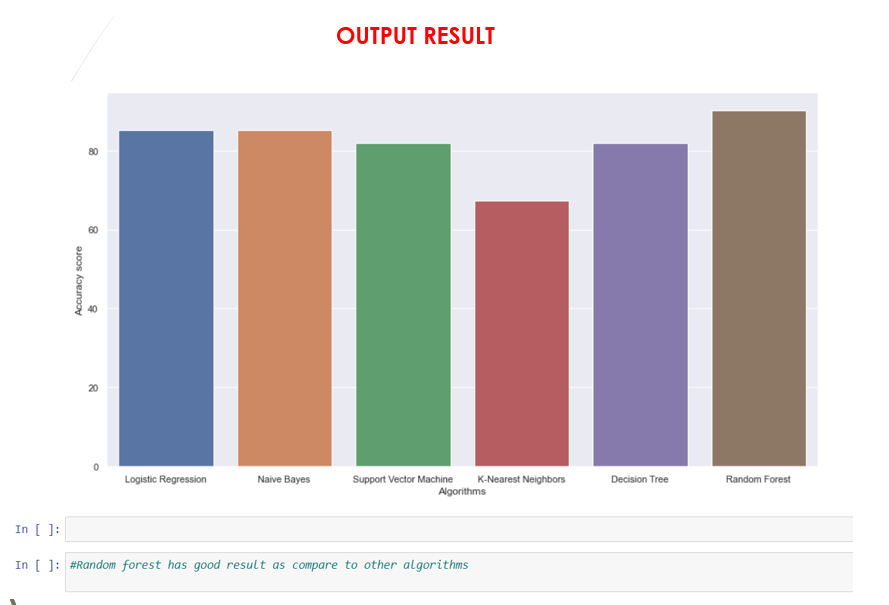
**Figure 12-SCREENSHOT-11**



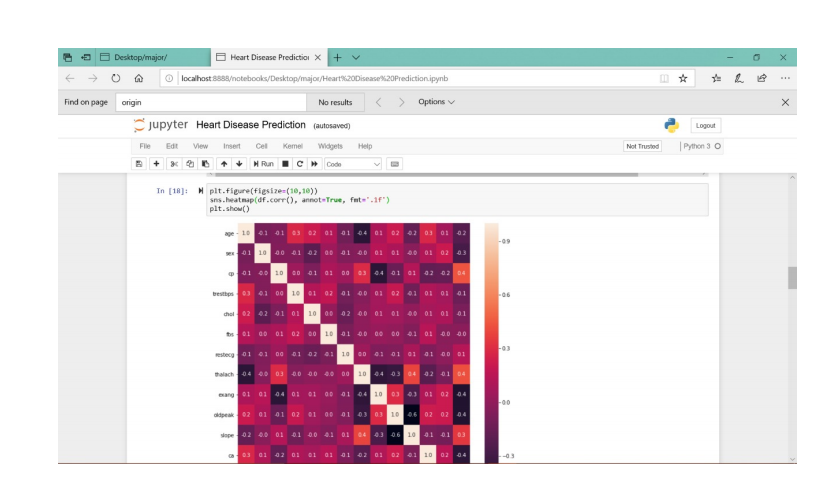
**Figure 13-SCREENSHOT-12**



**Figure 14-SCREENSHOT-13**



**Figure 15-SCREENSHOT-14**



**Figure 16-SCREENSHOT-15**

# 

# Chapter 7

# Conclusion and Future Scope

## 7.1 Conclusion

Finally our system can say with 86 per cent accuracy whether a person has any sort of heart ailments or not based on the details provided. Machine learning techniques were used in this work to process raw data and provide a new and novel discernment towards heart disease. Heart disease prediction is challenging and very important in the medical field. However, the mortality rate can be drastically controlled if the disease is detected at the early stages and preventative measures are adopted as soon as possible. Finally our system can say with 86 percent accuracy whether a person has any sort of heart ailments or not based on the details provided. RANDOM FOREST scored the best score of 90.16% of accuracy.

## 7.2 Future Scope

As the dataset which we were provided was only a fraction of the original our models

have a scope of performing even better with more data.

Also, since in heart disease treatment, a major part is the proper intake of nutrients

and less oil based fatty foods, we can improvise to improve this model by connecting it

with a dietary app which will provide a disease based, diet everyday, which will include

breakfast, lunch, evening snacks as well as dinner, keeping all sorts of food allergies that

the patient may have, in mind. This can prove to be quite beneficial for heart patients.

# Chapter 8

# REFRENCES

1. [1] Bo Jin ,Chao Che, Zhen Liu, Shulong Zhang, Xiaomeng Yin, And Xiaopeng Wei, “Predicting the Risk of Heart Failure With EHR Sequential Data Modeling” ,IEEE Access 2018.
2. [2] Aakash Chauhan , Aditya Jain , Purushottam Sharma , Vikas Deep, “Heart Disease Prediction using Evolutionary Rule Learning”, “International Conference on "Computational Intelligence and Communication Technology” (CICT 2018).
3. [3] Ashir Javeed, Shijie Zhou, Liao Yongjian, Iqbal Qasim, Adeeb Noor, Redhwan Nour4, Samad Wali And Abdul Basit , “An Intelligent Learning System based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection” , IEEE Access 2017.
4. [4] Senthilkumar Mohan, Chandrasegar Thirumalai, and Gautam Srivastava, “Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques”, IEEE Access 2019.
5. [5] K.Prasanna Lakshmi, Dr. C.R.K.Reddy, “Fast Rule-Based Heart Disease Prediction using Associative Classification Mining”, IEEE International Conference on Computer, Communication and Control (IC4-2015).
6. [6] M.Satish, D Sridhar, “Prediction of Heart Disease in Data Mining Technique”, International Journal of Computer Trends & Technology (IJCTT), 2015.
7. [7] Lokanath Sarangi, Mihir Narayan Mohanty, Srikanta Pattnaik, “An Intelligent Decision Support System for Cardiac Disease Detection”, IJCTA, International Press 2015.
8. [8] Boshra Bahrami, Mirsaeid Hosseini Shirvani, “Prediction and Diagnosis of Heart Disease by Data Mining Techniques”, Journal of Multidisciplinary Engineering Science and Technology (JMEST) ISSN: 3159-0040 Vol. 2 Issue 2, February–2015.
9. [9] Mamatha Alex P and Shaicy P Shaji,“Prediction and Diagnosis of Heart Disease Patients using Data Mining Technique”, International Conference on Communication and Signal Processing 2019.
10. [10] Senthilkumar Mohan, Chandrasegar Thirumalai, Gautam Srivastava ―Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques‖, Digital Object Identifier 10.1109/ACCESS.2019.2923707, IEEE Access, VOLUME 7, 2019 S.P. Bingulac, ―On the Compatibility of Adaptive Controllers,‖ Proc. Fourth Ann. Allerton Conf. Circuits and Systems Theory, pp. 8-16, 1994. (Conference proceedings)
11. [11] Sonam Nikhar, A.M. Karandikar” Prediction of Heart Disease Using Machine Learning Algorithms” International Journal of Advanced Engineering, Management and Science (IJAEMS) Infogain Publication,[Vol-2, Issue-6, June- 2016].I.S. Jacobs and C.P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.
12. [12] Aditi Gavhane, Gouthami Kokkula, Isha Pandya, Prof. Kailas Devadkar (PhD),” Prediction of Heart Disease Using Machine Learning”, Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018).IEEE Conference Record # 42487; IEEE Xplore ISBN:978-1- 5386-0965-1
13. [13] Abhay Kishore1, Ajay Kumar2, Karan Singh3, Maninder Punia4, Yogita Hambir5,” Heart Attack Prediction Using Deep Learning”, International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 04 | Apr-2018.
14. [14] A.Lakshmanarao, Y.Swathi, P.Sri Sai Sundareswar,” Machine Learning Techniques For Heart Disease Prediction”, International Journal Of Scientific & Technology Research Volume 8, Issue 11, November 2019.
15. [15] Mr.Santhana Krishnan.J, Dr.Geetha.S,” Prediction of Heart Disease Using Machine Learning Algorithms”,2019 1st International Conference on Innovations in Information and Communication Technology(ICIICT),doi:10.1109/ICIICT1.2019.8741465.
16. [16] Avinash Golande, Pavan Kumar T,” Heart Disease Prediction Using Effective Machine Learning Techniques”, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1S4, June 2019.
17. [17] V.V. Ramalingam, Ayantan Dandapath, M Karthik Raja,” Heart disease prediction using machine learning techniques: a survey”, International Journal of Engineering & Technology, 7 (2.8) (2018) 684-687
18. [18] Pranav Sorte, Avinash Golande, Advait Yermalkar, Vikas Suryawanshi, Utkarsha Wanjare, Sandip Satpute, “Smart Hospital for Heart Disease Prediction Using IOT”, International Journal of Innovative Technology and Exploring Engineering, ISSN: 2278-3075, Volume-8 Issue-9, July 2019.
19. [19] Shweta Gajbhiye, Bharati Vyas, Anshika Janbandhu, Shrushti Shrikhande, Komal Nagpure, Mrunali Agashe, “Heart Attack Early Prediction Using IOT”, IOSR Journal of Engineering, ISSN (e): 2250-3021, ISSN (p): 2278-8719, PP 50-54, 2019.
20. [20] Ponugumatla Kalyan, Mr. Gouri Shankar Sharma, “IOT Based Heart Function Monitoring and Heart Disease Prediction System”, International Journal for Science and Advance Research in Technology, ISSN [ONLINE]: 2395- 1052, Volume 3 Issue 12 – DECEMBER 2017